

ChatGPT Beyond English: Towards a Comprehensive Evaluation of Large Language Models in Multilingual Learning

Year: 2023 | Citations: 343 | Authors: Viet Dac Lai, Nghia Trung Ngo, Amir Pouran Ben Veyseh

Abstract

Over the last few years, large language models (LLMs) have emerged as the most important breakthroughs in natural language processing (NLP) that fundamentally transform research and developments in the field. ChatGPT represents one of the most exciting LLM systems developed recently to showcase impressive skills for language generation and highly attract public attention. Among various exciting applications discovered for ChatGPT in English, the model can process and generate texts for multiple languages due to its multilingual training data. Given the broad adoption of ChatGPT for English in different problems and areas, a natural question is whether ChatGPT can also be applied effectively for other languages or it is necessary to develop more language-specific technologies. The answer to this question requires a thorough evaluation of ChatGPT over multiple tasks with diverse languages and large datasets (i.e., beyond reported anecdotes), which is still missing or limited in current research. Our work aims to fill this gap for the evaluation of ChatGPT and similar LLMs to provide more comprehensive information for multilingual NLP applications. While this work will be an ongoing effort to include additional experiments in the future, our current paper evaluates ChatGPT on 7 different tasks, covering 37 diverse languages with high, medium, low, and extremely low resources. We also focus on the zero-shot learning setting for ChatGPT to improve reproducibility and better simulate the interactions of general users. Compared to the performance of previous models, our extensive experimental results demonstrate a worse performance of ChatGPT for different NLP tasks and languages, calling for further research to develop better models and understanding for multilingual learning.